AN APPLICATION OF ARTIFICIAL NEURAL NETWORK CLASSIFIER FOR MEDICAL DIAGNOSIS

AMJED KHALEEL IBRAHEEM

A thesis submitted in partial
fulfillment of requirement for the award of the
Degree of Master of Computer Science

Faculty of Computer Sciences and Information Technology
University Tun Hussein Onn Malaysia

JULY 2013

ABSTRACT

In recent year, various models have been proposed for medical diagnosis, which broadly can be classified into physical-based approaches and statistical-based approaches. Uncertainty and imprecision are the most important problems in medical diagnosis, other many problems in medical diagnostic domains need to be represented at varying degrees of diagnosis to be solved. Moreover, classification is very important in computer-aided medical diagnosis. In this respect, Artificial Neural Network (ANN) have been successfully applied and with no doubt, they provide the ability and potentials to diagnose the diseases. Therefore, this research focuses on using ANN to classify medical data. ANN model with two layers of tunable weights were used and trained using four different backpropagation algorithms while are the gradient descent(GD), gradient descent with momentum(GDM), gradient descent with adaptive learning rate(GDA) and gradient descent with momentum and adaptive learning rate(GDX). The network was used to classify three sets of medical data taken from UCI machine learning repository. The ability of all training algorithms tested and compared to each other on all datasets. Simulation results proved the ability of ANN for medical data classification with high accuracy and excellent performance and efficiency. This research provides the possibility of reduce costs and human resources. Increasing speed to find the results of medical analysis by using ANN also contributes in saving time for both physicians and patients.

ABSTRAK

Pada kebelakangan ini, pelbagai model telah dicadangkan untuk diagnosis perubatan, yang secara umumnya boleh dikelaskan kepada pendekatan berasaskan fizikal dan pendekatan berasaskan statistik. Ketidakpastian dan ketidaktepatan adalah masalah yang paling penting dalam diagnosis perubatan selain terdapat banyak masalah-masalah lain dalam bidang-bidang diagnostik perubatan perlu diwakili oleh pelbagai peringkat diagnosis yang perlu diselesaikan. Selain itu, klasifikasi adalah sangat penting dalam diagnosis perubatan dengan bantuan komputer. Oleh itu, dalam hal ini, rangkaian neural tiruan (ANN) telah berjaya digunakan dan dengan tidak syak lagi, mereka menyediakan potensi untuk mendiagnosis penyakit. Sementara itu, kajian ini memberi tumpuan dengan menggunakan ANN untuk mengklasifikasikan data perubatan. Model ANN dengan dua lapisan pemberat boleh tala telah digunakan dan dilatih menggunakan empat algoritma rambatan balik berbeza iaitu penurunan kecerunan (GD), penurunan kecerunan dengan momentum (GDM), penurunan kecerunan dengan kadar pembelajaran adaptif (GDA) dan penurunan kecerunan dengan momentum dan kadar pembelajaran adaptif (GDX). Rangkaian ini telah digunakan untuk mengklasifikasikan kepada tiga set data perubatan yang diambil dari UCI mesin pembelajaran repository. Keupayaan semua algoritma latihan diuji dan dibandingkan dengan satu sama lain pada semua dataset. Keputusan simulasi membuktikan keupayaan ANN untuk pengelasan data perubatan dengan ketepatan yang tinggi dan prestasi yang sangat. Kajian ini menyediakan kemungkinan mengurangkan kos dan sumber manusia dan meningkatkan kelajuan dalam mencari keputusan analisis perubatan dengan menggunakan ANN juga menyumbang dalam menjimatkan masa doktor dan pesakit.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

η	-	Learning Rate
α	-	Momentum
$\boldsymbol{\delta}$	-	The information error of the hidden nodes
w	-	Weight
ACC	-	Accuracy
$f(\sum)$	-	Activation Function
X	-	The value of vector X
minX	-	The minimum value of vector X
maxX	-	The maximum value of vector X
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
BP	-	Back Propagation
FNA	-	Fine Needle Aspirate
FNN	-	Fuzzy Neural Network
GD	-	Gradient Descent
GDA	-	Gradient descent with Adaptive learning rate
GDM	-	Gradient Descent with Momentum
GDX	-	Gradient Descent with Momentum and Adaptive Learning Rate
MLP	-	Multi-Layer Perceptron
MSE	-	Mean Squared Error
PSO	-	Particle Swarm Optimization
SLP	-	Single Layer Perceptron
SVM	-	Support Vector Machine

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

All around the world there are many people who suffer from many diseases and health problems. Most of those diseases are simple and easy to be detected and diagnosed by doctors, however, there are some that are dangerous, and at the same time it is very difficult to diagnose them, resulting in the difficulty to prescribe a proper treatment.

Thus, early detection of medical problems such as; breast cancer, hepatitis, Parkinson, etc. is important to increase the chance of a successful treatment (Aribarg *et al.,* 2009). Traditional manual data analysis has become inefficient and methods for efficient computer based analysis are essential. Medical information systems in modern hospitals and medical institutions have widened causing great difficulties in extracting useful information for decision support (Jayalakshmi & Santhakumaran, 2010).

On the other hand, in computer aided decision systems, computer algorithms are used to help a physician in diagnosing a patient (Mazurowski *et al*., 2008). Machine learning classification techniques provide support for the decision-making process in many areas of health care, including screening, diagnosis, prognosis, monitoring, therapy, survival analysis, and hospital management (Abdel-Aal, 2005).

At the same time, medical diagnosis can be viewed as a pattern classification problem based on a set of input features. The goals to classify a patient as having or not having a particular disorder. This data can be classified by using different classification algorithms such as; Support Vector Machine(SVM), NaIve Bayes, K-Nearest Neighbour, Decision Tree and Artificial Neural Networks (ANN) (Peter & Somasundaram, 2012).

The field of ANN consists of a large collection of models and techniques originally inspired by biological nervous systems such as the human brain. Neural networks are based around simplified models of biological neurons. ANN is one of the powerful Artificial Intelligence (AI) techniques that have the capability to learn and memorise a set of data and construct weight matrices to represent the learning patterns. ANN is a mathematical model which emulates the activity of biological neural networks in the human brain (Sharif *et al*., 2011). Therefore, this research relates how to apply ANN for medical diagnosis to dermatology, breast cancer and acute inflammation diseases.

1.2 Problem Statement

Uncertainty and imprecision are the most important problems in medical diagnosis as per the expert systems, and during past decades many researchers have been directed towards this area (Aribarg *et al.*, 2009; Mazurowski, *et al*., 2008; Neto & Barreto, 2009). Many problems in medical diagnostic domains need to be represented at varying degrees of diagnosis to be solved (Kim *et al*., 1992). Therefore, classification is very important in computer-aided medical diagnosis (Dong & Zhang, 2010).

Consequently, accuracy is very essential in classifiers used for medical applications. A high percentage of false negatives in screening systems increases the risk of real patients not receiving the attention they need, while a high false alarm rate causes unwarranted worries and increases the load on medical resources (Abdel-Aal, 2005).

Intelligent tools that are commonly being used for medical diagnosis include Bayesian and Nearest-Neighbour Classifiers, Rule Induction Methods, Decision Trees, Fuzzy Logic, SVM and ANN (Mazurowski *et al*., 2008). Among all these techniques, ANN allows easier model development and provides more transparency and greater insight into the modelled phenomena, which are important advantages in medicine (Abdel-Aal, 2005; Dancey *et al*., 2010; Mazurowski *et al*., 2008). ANN has important properties that are able provide solutions for complex classification problems, especially those of non-linear and high dimensionality (Neto & Barreto, 2009).

As ANN has some advantages such as, its nonparametric and non-linear nature, arbitrary decision boundary capabilities, easy adaptation to different types of data and input structures, and good generalisation capabilities, it has been successfully used in many applications including pattern classification, decision making, forecasting, computer-vision, and adaptive control (Neto & Barreto, 2009; Dancey *et al*., 2010; Mazurowski *et al*., 2008; Sharif *et al*., 2011; Jayalakshmi & Santhakumaran, 2010). In this study, in order to classify the medical data, the ANN technique is used as a data-driven model to alleviate the difficulty in diagnosing a certain disease.

1.3 Aim of the Study

The aim of this study is to classify medical data by using Artificial Neural Networks technique.

1.4 Objectives of the Study

In order to achieve the above mentioned research’s aim, a few objectives have been set, which are listed below:

i. To construct a classifier model by using Artificial Neural Networks (ANN).
ii. To train ANN using 4 different training algorithms.
iii. To classify medical data using the trained ANN model in (ii).
iv. To evaluate the performance of the trained ANN model.

1.5 Scope of Study

This research focuses on the use of ANN namely; the Multi-Layer Perceptron (MLP) as data classifier; trained with 4 different training algorithms: gradient descent (GD), gradient descent with momentum (GDM), gradient descent with adaptive learning rate (GDA) and gradient descent with momentum and adaptive learning rate. This study classified Dermatology data into six classes which are Psoriasis, Seboreic Dermatitis, Lichen Planus, PityriasisRosea, Cronic Dermatitis and Pityriasis Rubra Pilaris; Breast Cancer data into two classes which are Benign or Malignant; and Acute inflammation data into two classes which are Inflammation of Urinary Bladder or Nephritis of Renal Pelvis Origin. The data were taken from UCI machine learning web site. Results from the MLP for the classification task have been compared for different training algorithms.

1.6 Significance of Study

The importance of this research is to increase the classification accuracy by using ANN model into medical analysis. Therefore, it can improve repeated medical tests and neglect traditional manual analysis. Moreover, it can reduce costs and human resources. Increasing speed to find the results of medical analysis by using ANN also contributes in saving time for both physicians and patients.

1.7 Thesis Outline

The remaining part of this thesis is broken up into the following chapters. Chapter 2 is concerned with the relevant background information regarding using ANN for classification in the following order: (1) overview of ANN's architecture in terms of feed forward and feedback, (2) the advantages of ANN's (3) several techniques and applications that have been employed in classification medical data and (4) the learning algorithms that are used to train the network models. This chapter also highlights the virtues and limitations of each method and arguments are brought forward for alternative methods that can be used for classification medical data.

Chapter 3 describes brief steps on how to use the ANN models for classification of medical data, starting from the variable and data selection, data pre-processing and data partition, and performance comparison of the different training algorithms. The rationale of selecting parameters for each algorithm, the evaluation covering all the network parameters: the hidden nodes higher order terms, the learning factors and also the number of output nodes in the output layer.

The simulations for the comprehensive evaluation of the training algorithms are presented in Chapter 4. The simulation results of each training algorithm. Each model is then presented graphically. The last chapter, Chapter 5 concludes the work done and several recommendations are suggested in order to improve the performance of the proposed network models.

1.8 Summary

There are many applications and techniques on medical diagnosis that had been developed in the past. However, some limitations such as the accuracy and complexity of the models have made the existing system not so dependable for some applications. Therefore, diagnosis improvement requires continuous efforts in many fields, including ANN. Classification of medical data helps to detect and diagnose diseases. ANN classifiers have proven their ability to classify several types of data. This research focuses on using ANN on medical diagnoses and help to predict the diseases.

CHAPTER 2

LITERATURE REVIEW

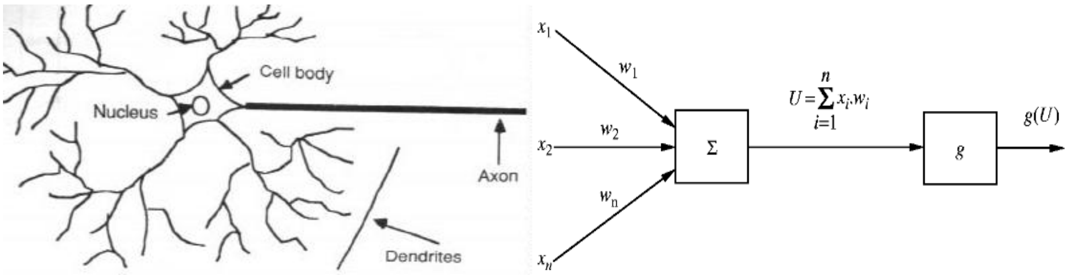
2.1 Introduction

In recent years, Artificial Intelligence (AI) research has been on the pioneering end that was increasingly being used in computer sciences, social sciences, physical sciences, engineering and statistics, with the purpose of modelling complex problems (Jones, 2008). AI deals with intellectual mechanisms which are the kind of abilities that a human uses to solve complex problems (Chen, 2005; Jones, 2008; Ham, & Kostanic, 2001). These intelligent mechanisms, which include Artificial Neural Network (ANN), offer great advantages over conventional modelling, including the neural structure of the brain that mimics the learning capability from experiences, and the ability to handle large amounts of noisy data from dynamic and nonlinear processes where nonlinearities and variable interactions play a vital role. Meanwhile, many research projects have shown that ANN is a powerful technique for several problems. Therefore, in order to be more certain in this field, this chapter provides the theoretical perspectives of a wide range of ANN which partly reveals the applications and techniques that have been used in ANN. This chapter also discusses the related works that are in line with the problem under study, the classification medical data.

2.2 From Biological Neuron to Artificial Neuron (Perceptron)

The human brain consists of a large number of neural cells, more than a billion that process information. Each cell works like a simple processor (James & Anderson 1997). A biological neuron consists of three main components, as shown in Figure 2.1(a): (i) dendrites that channel input signals, which are the connection strength, to a cell body. (ii) cell body(soma) which accumulates the weighted input signals and further processes these signals, and (iii) axon, which transmits the output signal to other neurons that are connected to it (Samarasighe, 2006).

Meanwhile, an artificial neuron consists of three main components, like a biological neuron, as shown in Figure 2.1(b): a set of input connections brings in activation from other neurons, a processing unit sums the inputs and then applies a non-linear activation function, and an output line transmits the result to other neurons.



(a): Biological Neuron (b): Artificial Neuron

Figure 2.1; From Biological Neuron to Artificial Neuron (Perceptron)

2.3 Artificial Neural Network

Artificial Neural Networks (ANN) are models that attempt to mimic some of the basic information processing methods found in the human brain. As our brains perform complex tasks, ANN modelled after the brain has also been found useful in solving complex problems (Samarasighe, 2006). Also an ANN is a massively parallel distributed processor that has a natural propensity for storing experimental knowledge and making it available for use (Dey *et al.*, 2008).

ANN consists of sets of input layer, hidden layer, output layer, each layer consists of a number of neurons and weighting functions (Naim *et al.*, 2011), as shown in Figure 2.2. The artificial neurons are organized in layers with one or more intermediate hidden layers put in between the input layer and output layer, and send their signals “forward”. Each layer has a number of neurons connected with neurons in the adjacent layers through unidirectional connections. The information flow is only allowed in one direction during the training process that is from the input layer to the output layer through the hidden layers. There can be any number of hidden layers in the architecture. The hidden layer has a synaptic weighting matrix and the weights are associated with all the connections made from the input layer to the hidden layer (Lee *et al.*, 2007; Junfeng & Leping, 2010; Dey *et al.,* 2008).

There are many different types of ANN, from relatively simple to very complex, just as there are many theories on how biological neural processing takes place (Abhishek *et al*, 2012). ANN performs a variety of tasks, including prediction or function approximation, pattern classification, clustering, and forecasting (He & Xu, 2009). Nevertheless, its performance is affected by how the setup of the neural networks structure is conducted, and by how data is prepared for it (Ogasawara *et al.*, 2009).

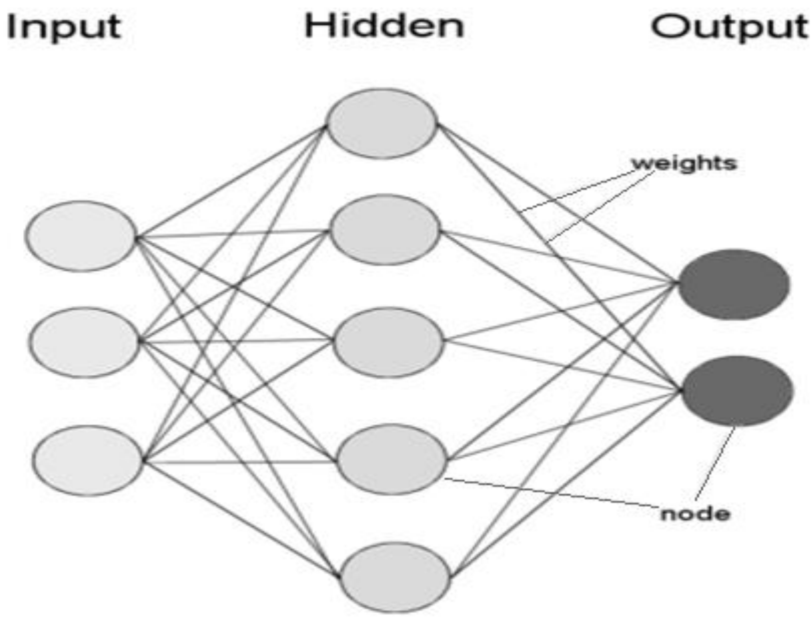


Figure 2.2: Artificial Neural Network (ANN)

2.3.1 Single Layer Perceptron (SLP)

An arrangement of one input layer of neurons is feed-forwarded to one output of neurons known as Single Layer Perceptron (SLP), as shown in Figure 2.3. SLP comprises of a single layer of weights, whereby the inputs are directly connected to the outputs, via a series of weights. The synaptic links carrying weights connect every input to every output, but not in the other way. This way is considered a network of feed-forward type (Martin, Howard & Mark 1996). This type of networks can perform pattern classification only on linearly separable patterns, regardless of the form of nonlinearity used. Linear separation requires that the pattern is classified to be sufficiently separated from each other to ensure that the decision boundaries are hyper planes (Yeung *et al.*, 2010).

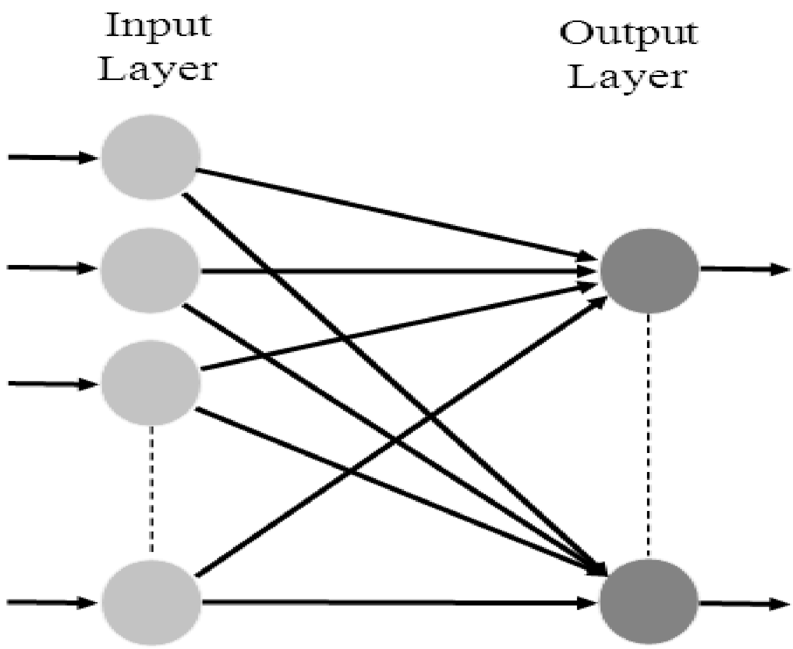


Figure 2.3: Single Layer Perceptron (SLP)

2.3.2 Multilayer Perceptron (MLP)

The architecture of this class of network, besides having the input and output layers, also have one or more intermediary layers called hidden layers, as shown in Figure 2.4. The computational units of the hidden layer are known as hidden neurons (Martin, Howard & Mark 1996). Each neuron is fully connected to all the neurons in its preceding layer as well as in all its next layers (Yeung *et al.*, 2010). MLP can solve more complicated problems than can SLP (Graupe, 2007). However, training might be more difficult. Nevertheless, in some cases, training may be more successful, because it is possible to solve a problem that SLP cannot be trained to perform correctly at all (Fausett, 1994).

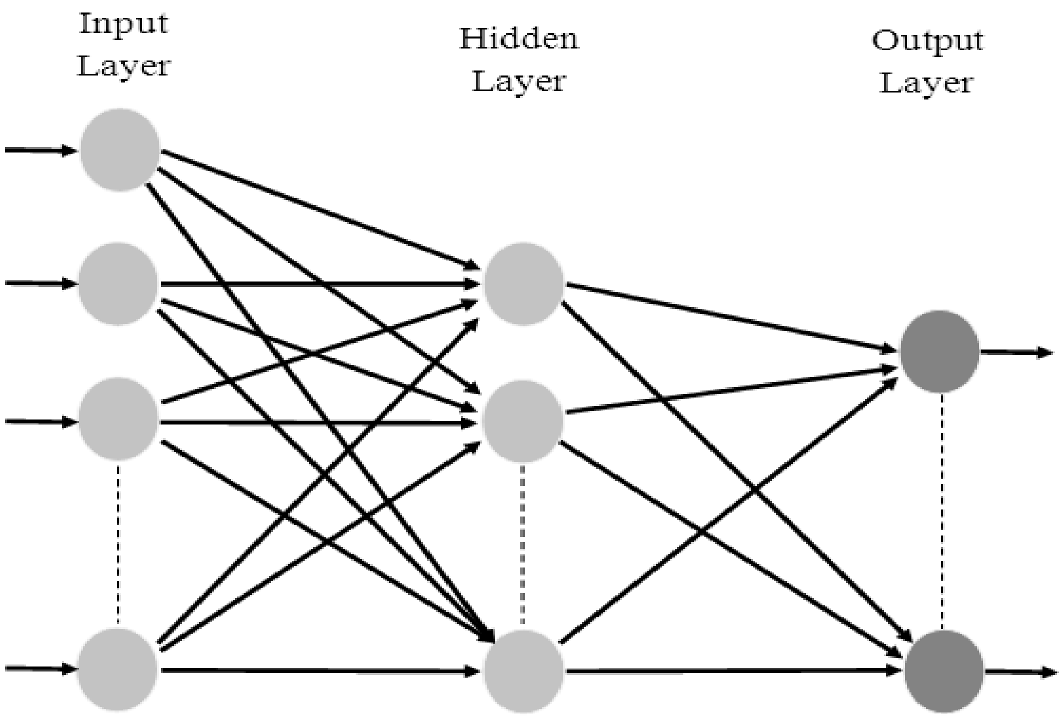


Figure 2.4: Multi-Layer Perceptron(MLP)

2.3.3 Backpropagation Algorithm

Backpropagation is the most popular of the MLP learning algorithms (Seiffertt & Wunsch 2010; Al-Allaf, 2011). The backpropagation algorithm can be defined as follows. For a test set, propagate one test through the MLP in order to calculate the output.

$$\boldsymbol{h_i = f \sum x_i\, w_{ij}} \tag{2.1}$$

$$\boldsymbol{y_i = f \sum h_i\, w_{jk}} \tag{2.2}$$

Where ***h*** is the hidden node
x is the input need
w is the weight
y is the output node

Then compute the error, which will be the difference of the expected value $\boldsymbol{t}$ and the actual value, and compute the error information term $\boldsymbol{\delta}$ for both of output and hidden nodes.

$$\boldsymbol{\delta y_i = y_i\ (1 - y_i)\ .\ (t - y_i)} \quad (2.3)$$

$$\boldsymbol{\delta h_i = h_i\ (1 - h_i)\ .\ \delta y_i\ .\ w_{jk}} \quad (2.4)$$

$\boldsymbol{\delta_j}$ the information error of the nodes

Finally, backpropagate this error through the network by adjusting all of the weights; starting from the weights to the output layer and ending at the weights to the input layer, as shown in Figure 2.5.

$$\Delta \boldsymbol{w_{jk} = \eta\ .\ \delta y_i\ .\ h_i} \quad (2.5)$$

$$\Delta \boldsymbol{w_{ij} = \eta\ .\ \delta h_i\ .\ x_i} \quad (2.6)$$

$$\boldsymbol{w_{new}} = \Delta \boldsymbol{w} + \boldsymbol{w_{old}} \quad (2.7)$$

Where $\boldsymbol{\eta}$ is the learning rate

Backpropagation adjusts the weights in an amount proportional to the error for the given unit (hidden or output) multiplied by the weight and its input. The training process continues until some termination criterion, such as a predefined mean-squared error, or a maximum number of epochs.

The backpropagation algorithm is a typical supervised learning algorithm, where inputs are provided and propagated forward to generate one or more outputs. Given the output, the error is calculated using the expected output. The error is then used to adjust the weights. There are two types of error functions for backpropagation. The first error function is used for output cells, and the second is used on for hidden cells.

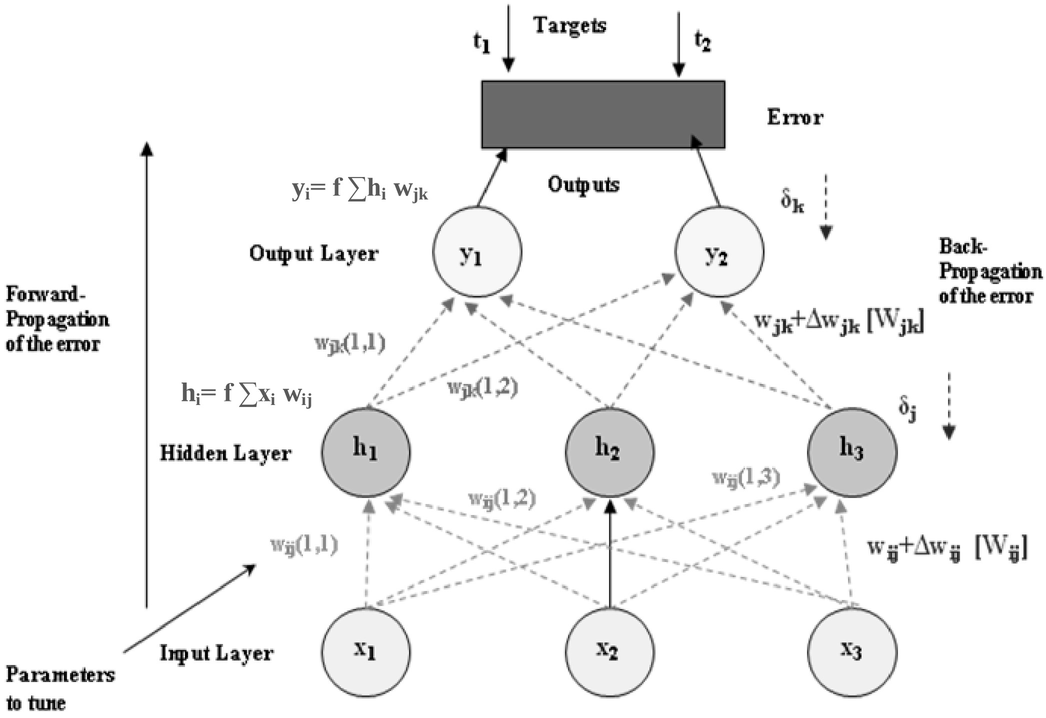


Figure 2.5: Backpropagation Neural Network

2.4 Component of Neural Network

A neural network is made up of a large number of simple processing elements called neurons, units, cells, or nodes. Each neuron is linked to other neurons by the techniques of directed communication links, each with the associated weight. The weights illustrate information being used by the net to solve a problem. Each neuron has an internal state, called its activation function, which is a function of the inputs that it has received. Normally, a neuron sends its activation as a signal to several other neurons. It is important to note that a neuron can send only one signal at a time, although that signal is broadcasted to several other neurons (Fausett, 1994; Padhy, 2005).

2.4.1 Neuron

The fundamental processing element of a neural network is a neuron. An artificial neuron is a mathematical function conceived as a simple model of a real (biological) neuron. Neurons are the basic computing unit that performs on the local data being processed inside a network. These neurons form massively parallel networks, whose function is determined by the network structure (Samarasighe, 2006). A neuron is often called a node or unit. It receives input from some other nodes by inputs, or perhaps from an external source. Each input has an associated weight (*w)*. The node computes some function (*f)* of the weighted sum of its inputs. Its output, in turn, can serve as input to other units.

2.4.2 Weight

ANN consists of two or more layers, each layer contains a number of nodes (neurons), as shown as Figure 2.5. The neurons are joined by directed arcs-connections. The neurons and arcs constitute the network topology. Each arc has numerical weight that specifies the influence between two neurons. Positive weights indicate reinforcement; negative weights represent inhibition (Yeung *et al.*, 2010).

2.4.3 Activation Function

Over the years, researches have tried several functions to convert the input into an output (Fausett, 1994). The neural activation functions have some important characteristics that make them vital to neural information processing. They are nonlinear, continuous functions that remain within some upper and lower bounds. Nonlinear means that the output of the function varies nonlinearly with the input (Samarasighe, 2006).

Activation function does the final mapping of the activation of the output neurons into the network outputs. But, the outputs from a single cycle of the operation of neural network might not be the final outputs, as the network has to comply with the convergence criterion (Samarasighe, 2006). The output of any neuron is thus a result of thresholding or its internal activation, if any. Activation function is important for multi-layer networks to preserve meaningful range areas of each layer’s operations (Rao & Srinivas, 2003).

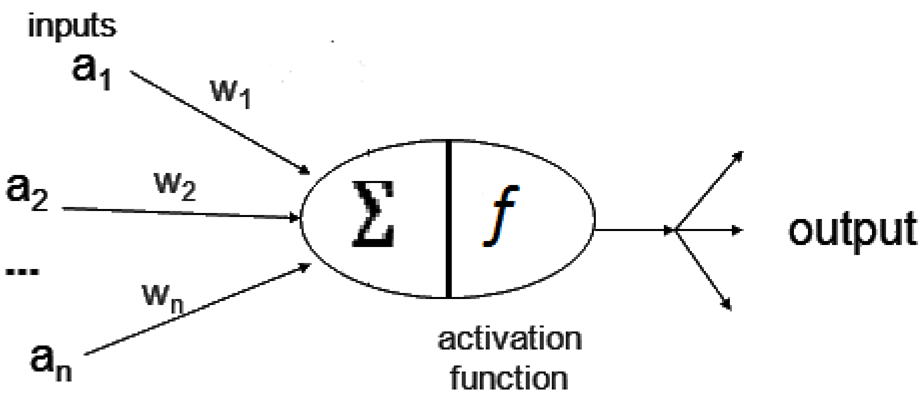


Figure 2.6: Activation function

The activation function, denoted by $f(\sum)$, as shown in Figure 2.5, defines the output of a neuron in terms of the induced local field $\sum$ (Haykin, 1999). There are three basic types of activation function:

2.4.3.1 Threshold Function

A threshold (hard-limiter) activation function can either a binary type or a bipolar type, as shown in Figure 2.6(a). Neuron with hard-limiter activation function is known as the McCulloch-Pitts model.

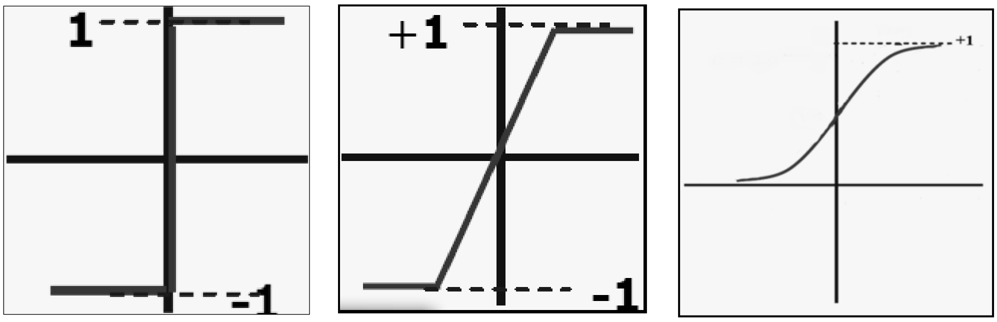
2.4.3.2 Piecewise Linear Function

This activation function is also known as saturating linear function and can have either a binary or bipolar range for the saturation limits of the output. The mathematical model representing a symmetric saturation function is shown in Figure 2.6(b).

2.4.3.3 Sigmoidal Function (S-shape function)

The nonlinear curved S-shape function is commonly known as the sigmoid function, as shown in Figure 2.6(c). This is the most widely type of activation used to construct the neural networks. It performs mathematically well with a differentiable and a strictly increasing function.

The sigmoidal function is reached by using the exponential equation. It can be obtained by changing the different shapes of the function, which adjusts to the abruptness of the function as it changes between the two asymptotic values.



(a) Threshold Function. (b) Piecewise Linear Function. (c) Sigmoidal Function

Figure 2.7: Types of Activation Function

2.5 Learning Algorithm for ANN

The learning (or training) for a NN is not simply a matter of memorizing the mapping relationship between the inputs and outputs among the learning samples, but of extracting the internal rules about the environment which are hidden in the sample by learning the finite sample data.

Up to now, there are many learning algorithm of neural networks among which are error back-propagation algorithm (BP algorithm) and its various improved patterns are most extensively and effectively applied.

MLP model which adopts the BP algorithm is generally called a BP network, and its learning process is made up of two parts: forward-propagation of input information and error back-propagation. Forward-propagation input information is transferred to the output layer from the input layer after being processed in the hidden layer. The state of each layer neuron influences only the state of neurons in the next layer. If it does not obtain the expected output in the output layer, it shifts to back-propagation, and error signals are shown along the original pathway of the neural connection, in return the connection weight of each layer is modified one by one. Through successive iterations, the error between the expected output signals of the network and practical output signals of the system reaches an allowable range.

A learning algorithm for an ANN is often related to a certain function approximation algorithms, especially to some iterative algorithms that make the approximation error gradually smaller. In fact, the above-mentioned BP algorithm corresponds to gradient descent algorithms; such as gradient descent, gradient descent with momentum and gradient descent with adaptive learning rate in function approximation. Once this principle is known, it can construct various learning algorithms for neural networks according to different function approximation algorithms.

2.5.1 Gradient Descent Backpropagation (GD)

GD is a network training function that informs about weight and bias values according to gradient descent. The weights and biases are kept abreast in the direction of the negative gradient of the performance function.

The learning rate η is multiplied by the negative of the gradient to conclude the changes to the weights and biases, as obtained in Equation 2.8.

$$\Delta w_{ij} = \eta \cdot \delta_j \cdot x_{ij} \tag{2.8}$$

Where Δw_{ij} is the delta/gradient of weights
η the learning rate parameter
δ_j the information error of the nodes
And x_{ij} the value of the network nodes

The larger the learning rate, the bigger the step. If the learning rate becomes too large, the algorithm will be unstable. If the learning rate is fixed too small, the algorithm will take a long time to converge. Apart from GD, there are three other variations of gradient descent.

2.5.2 Gradient Descent with Momentum Backpropagation (GDM)

GDM allows a network not only to respond to the local gradient, but also to track recent trends in the error surface. It acts like a low pass filter, a momentum α which allows the network to disregard small features in the error surface. Without momentum a network can get stuck whereas, with momentum a network can slide through such a minimum (Li *et al.*, 2009; Kathirvalavakumar & Subavathi, 2012).

GDM depends on two training parameters. The parameter learning rate is similar to the simple gradient descent. The parameter momentum is the momentum constant that defines the amount of momentum, as in Equation 2.9.

$$\Delta w_{ij}(r) = \eta \, . \, \delta_j \, . \, x_{ij} + \alpha \, . \, \Delta w_{ij}(r-1) \qquad (2.9)$$

Where α is the momentum parameter
And r the of iteration

2.5.3 Gradient Descent with Adaptive Learning Rate (GDA)

GDA is a network training function that brings up-to-date weight and bias values as per the gradient descent with adaptive learning rate. With standard steepest descent, the learning rate remains stable throughout the training. The performance of the algorithm is directly related to the proper setting of the learning rate. If the learning rate is set too high, the algorithm can vacillate and become unstable. If the learning rate is too small, the algorithm will take a long time to converge (Chuan, 2010; Sheel1 & Varshney, 2007). The optimal setting for the learning rate cannot be determined before training and in fact, the optimal learning rate varies during the training process, as the algorithm moves across the performance surface. It can improve the performance of the steepest descent algorithm if the learning rate is allowed to change during the training process. An adaptive learning rate attempts to keep the learning step size as large as possible while keeping learning stable. The learning rate is directly related to the complexity of the local error surface (Chuan, 2010).

An adaptive learning rate necessitates some changes in the training procedure used by GDA. Firstly, the initial network output and error are calculated. At each epoch new weights and biases are calculated by using the current learning rate. New outputs and errors are then calculated. Equation 2.10 were used for calculate the adaptive learning rate.

$$\eta_{ij}\,(t) = \eta_{ij}\,(t-1) + \Delta w_{ij}(t) \, . \, \Delta w_{ij}(t-1) \qquad (2.10)$$

2.5.4 Gradient Descent with Momentum and Adaptive Learning Rate (GDX)

The function GDX combines adaptive learning rate with momentum training. It is done in the same way as GDA, except that it has the momentum coefficient as an additional training parameter, as in Equations 2.9 and 2.10.

2.6 Classification Using ANN

The ANN is an important and successful application because of the characteristics of its information-processing mechanism (He & Xu, 2009), and it has been successfully applied to broad spectrum of data-intensive applications.

ANNs have gained popularity in various fields due to its ability to solve highly complex problems such as; classification, detection, and prediction with good accuracy (Law *et al*., 2006; Naim *et al*., 2011; Krogh, 2008). NNs have found a profound success in the area of classification. By repeatedly showing a neural network inputs classification into groups, the network can be trained to discern the criteria used to classify, and it can do so in a generalized manner allowing successful classification of new inputs not having been used during training (Siraj *et al.*, 2006).

Over the years, significant researches have been performed for different applications of classifications. One of these researches, Naim *et al*., (2011) proposed the classification system for fingerprint images using NN with good performance and high accuracy. In 2010, Zhu *et al*, tried to classify the bacteria. The results proved that ANN is feasible and efficient. Han, Cheng & Meng (2002), proposed the Classification of Aerial Photograph Using Neural Network. The results demonstrated that the neural network suits the classification of remotely sensed data and is superior to the maximum likelihood classifier in the accuracy of the classification and has an overall effect. Daud *et al* (2011), has been classified as spoken letters by using neural network. The results indicate a classification accuracy of 100% (training) and 93% (testing). From these works, the capability of neural network is proved in the field of classification and its high performance.

2.7 Classification of Medical Data

Computer Aided Diagnosis is based on classification of medical data by intelligent classifiers. Especially for medical purposes, the classification must be very efficient, as diagnosis demands a high rate of reliability (Tsirogiannis *et al*., 2004).

In the past, ANN techniques have been used to solve many classification problems of medical data and contribute to medical diagnosis. One of the researches related to the use of ANN in medical diagnosis was done by Jayalakshmi & Santhakumaran (2010). Besides improving the classification accuracy, results from their research confirmed the capability of NNs in the classification problem when applied to diabetes data. Another interesting research which used NN as a classifier in diagnosing tuberculosis was carried out by Rulaningtyas *et al*., (2011) and the results were successful in reducing classification error.

Meanwhile, (Mazurowski *et al*., 2008; Aribarg *et al.,* 2009), proposed a classification computation which referred to a modified NN with Particle Swarm Optimization (PSO). The results showed that classifier performance deteriorates with even a modest class imbalance in the training data. Further, it was shown that NN is generally preferable over PSO for imbalanced training data especially, with small data sample and large number of features.

Another research was done by using Fuzzy Neural Network (FNN) for cancer classification (Chu *et al*., 2004). FNN classifier helps biological researchers to differentiate cancers that are difficult to be classified using traditional clinical methods. Meanwhile, Trung *et al.* (2010), used SVM to classify medical imaging into two classes; benign and malignant for breast cancer treatment. However, this method did not reduce the noisy data and outliers (Trung *et al.*, 2010). Most of the aforementioned researches pointed the power and robustness of ANN. However, the complexities of architecture and time are still nontrivial difficulties. This study, thereby, presents the application of ANN to classify medical data.

2.8 Chapter Summary

In medical diagnosis, a few things such as; the techniques and approaches of dealing with medical data need to be considered. Therefore, in this chapter, several methodologies, applications and techniques that lie between empirical and physical studies have been discussed to improve the proper medical diagnosis capability for diseases. Likewise, this chapter also highlights the ANN approaches with the use of BP learning algorithm. Moreover, this chapter also discusses four types of training algorithms. The next chapter indicates the research framework that will be used to practically test the theoretical perspectives that have been described in this chapter.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This section provides detailed properties of the proposed neural network model, and brief steps on how to design the network model for medical diagnosis. The methodology used in this study briefly involves the following steps, as shown in Figure 3.1: (1) the first step is data selection which will be deliberated in the network simulation, (2) the pre-processing procedure which is used to remove the inconsistencies of the data, (3) the data partition that segregates the data into 2 parts; training and testing, (4) network model topology which maps the interconnection between nodes and selection of network structure, which entail the number of input-output nodes, hidden nodes and transfer function, (5) training of the network which considers the experiments being made in order to validate the proposed network model by using the back-propagation algorithm, and (6) model selection which entails the method of model selection based on overall performance. The unifying goal of this research method is to discover the network model which achieves diagnostic accuracy.

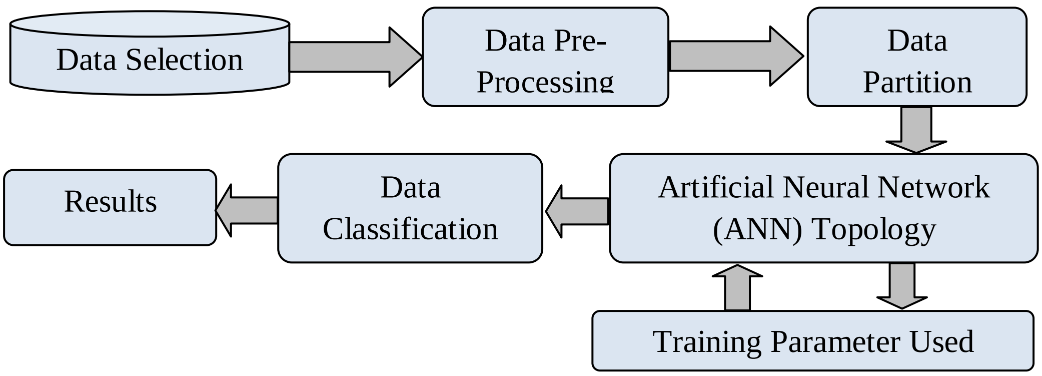


Figure 3.1: Methodology Framework

3.2 Data Selection

In this work, three different types of data specialized in the medical field were selected, taken from UCI machine learning repository (Hopper *et al.*, 1999). The first data set were collected from patients suffering of Dermatology. The second data set were collected from patients having Breast Cancer. The third one, were collected from patients suffering from Acute Inflammations in the urinary system. These data were chosen because they represent a sufficient pool of diseases, which could help in testing the proposed system.

Each set of data includes a number of instances, a number of attributes and a number of classes. Table 3.1 shows the data description used in this research work.

Table 3.1: Content of Datasets

	Number of instances	**Number of attribute**	**classes**
Dermatology	366	34	6
Breast cancer	699	9	2
Acute inflammation	120	6	2

3.2.1 Dermatology Dataset

The differential diagnosis of erythemato-squamous diseases is a crucial problem in dermatology. They all share the clinical features of erythema and scaling, with almost no difference. Usually, a biopsy is necessary for the diagnosis but, unfortunately these diseases share many histopathological features as well. Another difficulty for the differential diagnosis is that a disease may show the features of another disease at the beginning stage and will show its own characteristic features in the next stages. Patients were first evaluated clinically with 12 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features, as shown in Table 3.2. The values of the histopathological features were determined by an analysis of the samples under a microscope, and diagnosed the diseases, as shown in Table 3.3.

Table 3.2: Attribute Information of Dermatology Dataset

Dermatology			
Attribute Information			
1	Erythema	2	Scaling
3	Definite Borders	4	Itching
5	Koebner Phenomenon	6	Polygonal Papules
7	Follicular Papules	8	Oral Mucosal Involvement
9	Knee And Elbow Involvement	10	Scalp Involvement
11	Family History	12	Melanin Incontinence
13	Eosinophils In The Infiltrate	14	PNL Infiltrate
15	Fibrosis Of The Papillary Dermis	16	Exocytosis
17	Acanthosis	18	Hyperkeratosis
19	Parakeratosis	20	Clubbing Of The Rete Ridges
21	Elongation Of The Rete Ridges	22	Thinning of The Suprapapillary Epidermis
23	Spongiform Pustule	24	Munro Microabcess
25	Focal Hypergranulosis	26	Disappearance Of The Granular Layer
27	Vacuolisation And Damage Of Basal Layer	28	Spongiosis
29	Saw-Tooth Appearance Of Retes	30	Follicular Horn Plug
31	PerifollicularParakeratosis	32	Inflammatory MonoluclearInflitrate
33	Band-Like Infiltrate	34	Age

Table 3.3: Class Information of Dermatology Dataset

Dermatology		
Class Information		
1	Psoriasis	112 Instances
2	Seboreic Dermatitis	61 Instances
3	Lichen Planus	72 Instances
4	PityriasisRosea	49 Instances
5	Cronic Dermatitis	52 Instances
6	PityriasisRubraPilaris	20 Instances

3.2.2 Breast Cancer Dataset

Breast cancer dataset features were computed from a digitized image of a Fine Needle Aspirate (FNA) of a breast mass. They define characteristics of the cell nuclei present in the image. Nine real-valued features were computed for each cell nucleus, as shown in Table 3.4. And the diseases were limited to two classes; benign or malignant, as shown in Table 3.5.

Table 3.4: Attribute Information of Breast Cancer Dataset

Breast Cancer			
Attribute Information			
1	Clump Thickness	2	Uniformity of Cell Size
3	Uniformity of Cell Shape	4	Marginal Adhesion
5	single Epithelial Cell Size	6	Bare Nuclei
7	Bland Chromatin	8	Normal Nucleoli
9	Mitoses		

Table 3.5: Class Information of Breast Cancer Dataset

Breast Cancer		
Class Information		
1	Benign	458 Instances
2	Malignant	241 Instances

3.2.3 Acute Inflammation Dataset

This data set represents two diseases of the urinary system, the acute inflammations of urinary bladder and acute nephritis. Acute inflammation of the urinary bladder is characterised by the sudden occurrence of pains in the abdomen region and the urination in the form of constant urine pushing, micturition pains and sometimes lack of urine. Temperature of the body is rising however, most often not above 38^0C. The excreted urine is turbid and sometimes bloody. With proper treatment, symptoms decay usually within several days. However, there is an inclination for the urinary

REFERENCES

Abdel-Aal, R. E. (2005). *Improved Classification of Medical Data Using Abductive Network Committees Trained on Different Feature Subsets*. King Fahd University of Petroleum and Minerals: Ph.D. Thesis.

Abhishek, K., Kumar, A., Ranjan, R. & Kumar, S. (2012). A Rainfall Prediction Model using Artificial Neural Network. *Control and System Graduate Research Colloquium (ICSGRC),* pp. 82 – 87.

Al-Allaf, O.N.A. (2011). Fast BackPropagation Neural Network Algorithm for Reducing Convergence Time of BPNN Image Compression. *International Conference on Information Technology and Multimedia (ICIM),* pp. 1 – 6.

Aribarg, T., Supratid, S. & Chidchanok. L. (2009). Contemporary Classification on Medical Data based on Non-Linear Feature Extraction. *IEEE Int. Conf. on Computational Science and Its Application*, pp. 17 – 23

Chen, L. (2005). Pattern Classification by Assembling Small Neural Networks. *IEEE International Joint Conference on Neural Networks (IJCNN), vol(3),* pp. 1947 – 1952.

Chu, F., Xie, W. & Wang, L. (2004). Gene Selection and Cancer Classification Using a Fuzzy Neural Network. *IEEE Annual Meeting of the Fuzzy Information, Processing NAFIPS '04, vol(2),* pp. 555 – 559.

Chuan, H. (2012). Height Conversion in Momentum and Adaptive Learning Rate Algorithm. *International Conference on Computer, Mechatronics, Control and Electronic Engineering (CMCE), 2010, vol(4),* pp. 92 – 95.

Curteanu, S., Leon, F., Furtuna, R., Dragoi, E. N. & Curteanu, N. (2010). Comparison between Different Methods for Developing Neural Network Topology Applied to a Complex Polymerization Process. *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pp. 1 – 8.

Dancey, D., Zuhair, A. B. & David, M. (2010). Rule Extraction from Neural Networks for Medical Domains. *IEEE Int. Joint Conf. Neural Network (IJCNN)*, pp.1 – 8.

Daud, M. S., Yassin, I. M., Zabidi, A., Johari, M. A. & Salleh, M.K.M. (2011). Investigation of MFCC Feature Representation for Classification of Spoken Letters using Multi-Layer Perceptrons (MLP). *International Conference on Computer Applications and Industrial Electronics (ICCAIE)*, pp. 16 – 20.

Dey, R.; Bajpai, V.; Gandhi, G. & Dey, B. (2008). Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus. *IEEE Region 10 and the Third Int. Conf. on Industrial and Information Systems (ICIIS)*, pp. 1 – 4.

Fausett, L. (1994). *Fundamentals of Neural Networks: Architectures, Algorithms and Applications.* Florida Institute of Technology: Prentice Hall.

Fletcher, L., Katkovnik, V., Steffens, F.E. & Engelbrecht, A.P. (1998). Optimizing the Number of Hidden Nodes of a Feedforward Artificial Neural Network. *The 1998 IEEE International Joint Conference on Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence, vol(2)*, pp. 1608 – 1612.

Graupe, D. (2007). *Principles of Artificial Neural Networks*. 2^{nd} ed. Chicago, USA: University of Illinois.

Ham, F. & Kostanic, I. (2001). *Principles of Neurocomputing for Science and Engineering*. McGraw Hill, New York, NY.

Han, M., Cheng, L. & Meng, H. (2002). Classification of Aerial Photograph Using Neural Network. *IEEE International Conference on Systems, Man and Cybernetics, vol (6)*, pp. 6 – 9.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. 2^{nd} ed. Hamilton, Ontario, Canada: McMaster University.

He, X. & Xu, S. (2009). *Process Neural Networks: Theory and Application*. China: Zhejiang University Press: Prentice Hall.

Hopper, A., Russell, A., Zhao, B. Y., Abiteboul, S., Milo, T., Blum, A. & Jennifer Widom, J. (1999). information retrieval & language model. *Monte Carlo*. Retrieved November 23, 2012, from UCI Machine Learning Repository

James, A. & Anderson, (1997). *An Introduction to Neural Network*. New Delhi: Prentice Hall.

Jayalakshmi, T. & Santhakumaran, A. (2010). A Novel Classification Method for Diagnosis of Diabetes Mellitus Using Artificial Neural Networks. *IEEE Int. Conf. on Data Storage and Data Engineering*, pp. 159 – 163.

Jones, M. T. (2008). *Artificial Intelligence A Systems Approach*. Infinity Science Press.

Junfeng, Z. & Leping, X. (2010). The Fault Diagnosis of Marine Engine Cooling System Based on Artificial Neural Network (ANN). *IEEE The 2nd Int. Conf. on Computer and Automation Engineering (ICCAE)*, pp. 186 – 189.

Karlik, B. & Olgac, A. V. (2007). Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks. *International Journal of Artificial Intelligence and Expert Systems (IJAE), vol(1), Issue(4),* pp. 111 – 122.

Kathirvalavakumar, T. & Subavathi, S. J. (2012). Modified Backpropagation Algorithm with Adaptive Learning Rate Based On Differential Errors and Differential Functional Constraints. *International Conference on Informatics and Medical Engineering (PRIME), (2012),* pp. 61 – 67.

Krogh, A. (2008). What are Artificial Neural Networks?. *Nature Biotechnology, vol. 26(2),* pp. 195 – 197.

Law, R. C., Cheang, R., Tan, Y. W. & Azid, I. A. (2007). Thermal Performance Prediction of QFN Packages using Artificial Neural Network (ANN). *IEEE 31st Int. Conf. on Electronics Manufacturing and Technology (ICEMT),* pp. 50 – 54.

Lee, H. W., Syono, M. I. & Azid, I. H. A. (2007). Application of Artificial Neural Network (ANN) for Predicting The Behaviour of Micromachined Diaphragm Actuated Electrostatically. *IEEE Int. Conf SENSORS (ICSENS),* pp. 316 – 319.

Li-dong, F. & Zhang Y. (2010). Medical Image Retrieval and Classification Based on Morphological Shape Feature. *3rd Int. Conf. on Intelligent Networks and Intelligent Systems,* pp. 116 – 119.

Li, Y., Fu, Y., Li, H. & Si-Wen, Z. (2009). The Improved Training Algorithm of Back Propagation Neural Network with Selfadaptive Learning Rate. *International*

Conference on Computational Intelligence and Natural Computing, 2009. (CINC), '09, vol (1), pp. 73 – 76.

Martin, H. T., Howard, D. H. & Mark, B. (1996). *Neural network design*. Boston: PWS Publ, Company.

Maudal, O. (1996). *Preprocessing data for Neural Network based ClassifiersRough Sets vs Principal Component Analysis*. University of Edinburgh: MSc. Thesis.

Mazurowski, M. A., Habas, P. A., Zurada, J. M., Joseph, Y. L., Baker, J. A. & Tourassi, G. D. (2008). Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance. *Elsevier Neural Networks 21*, pp. 427 – 436.

Naim, N. F., Yassin, A. I. M., Zakaria, N. & Ab Wahab, N. (2011). Classification of Thumbprint using Artificial Neural Network (ANN). *Int. Conf. on System Engineering and Technology (ICSET),* pp. 321 – 234.

Neto, A. R. R. & Barreto, G. A. (2009). On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis. *IEEE Latin America Transactions, vol. 7(4),* pp. 487 – 496.

Ogasawara, E., Murta, L., Zimbrao, G. & Mattoso, M. (2009). Neural Networks Cartridges for Data Mining on Time Series. *International Joint Conference on Neural Networks, IJCONN*, pp. 2302 – 2309.

Padhy, N. P. (2005). *Artificial Intelligence and Intelligent Systems*. Oxford University Press.

Peter, T. J. & Somasundaram, K. (2012). An Empirical Study on Prediction Of Heart Disease Using Classification Data Mining Techniques. *IEEE Int. Conf. on Advances in Engineering, Science and Management*, pp. 514 – 518.

Rao, M. A. & Srinivas, J. (2003). *Neural Networks: algorithms and Applications*. Pangbourne England: Alpha Science International.

Rulaningtyas, R., Suksmono, B. A. & Mengko, T. L. R. (2011). Automatic Classification of Tuberculosis Bacteria using Neural Network. *IEEE Int. Conf. on Electrical Engineering and Informatics,* pp. 1 – 4.

Samarasighe, S. (2006). *Neural Network for Applied Science and Engineering: From Fundamentals to Complex Pattern Recognition*. New York: Auerbach Publication.

Schalkoff, R. J. (1997). *Artificial Neural Networks*. New York: McGraw-Hill, Clemson University.

Seiffertt, J., Wunsch, D.C. (2010). Backpropagation and Ordered Derivatives in the Time Scales Calculus. *IEEE Transactions on Neural Networks, vol(21),* pp. 1262 – 1269.

Sharif, M. S., Abbod, M., Krill, B., Amira, A. & Zaidi H. (2011). Automatic PET Volume Analysis and Classification Based on ANN and BIC. *IEEE 15th International Symposium on Consumer Electronics*, pp. 565 – 570.

Sheel1, S., Varshney, T. & Varshney, R. (2007). Accelerated Learning in MLP using Adaptive Learning Rate with Momentum Coefficient. *International Conference on Industrial and Information Systems (ICIIS), (2007),* pp. 307 – 310.

Siraj, F., Yusoff, N. & Kee, L. C. (2006). Emotion Classification Using Neural Network. *International Conference on Computing & Informatics, ICOCI '06,* pp. 1 – 7.

Taylor, C. M. & Agah, A. (2006). Evolving Neural Network Topologies for Object Recognition. *Automation Congress, 2006. WAC '06. World*, pp. 1 – 7.

Trung L., Tran, D., Wanli, M. & Sharma, D. (2010). A New Support Vector Machine Method for Medical Image Classification. *IEEE European Workshop on Visual Information Processing (EUVIP),* pp. 165 – 170.

Tsirogiannis, G. L., Frossyniotis, D., Stoitsis, J., Golemati, S., Stafylopatis, A. & Nikita, K. S. (2004). Classification of Medical Data with a Robust Multi-Level Combination Scheme. *IEEE Int. Joint Conf. on Neural Networks, vol.* 3, pp. 2483 – 2487.

Utomo, W. M., Haron, Z. A., Bakar, A. A., Ahmad, M. Z. & Taufik. (2010). Voltage Tracking of a DC-DC Buck-Boost Converter using Neural Network Control. *International Journal of Computer Technology and Electronics Engineering (IJCTEE), vol(1), Issue(3),* pp. 108 – 113.

Virili, F. & Freisleben, B. (2000). Nonstationarity and Data Preprocessing for Neural Network Predictions of an Economic Time Series. *Proceedings of the IEEE-*

INNS-ENNS International Joint Conference on Neural Networks, IJCNN, Vol(5), pp. 129 – 134.

Vucetic, S., Fiez, T. & Obradovic, Z. (1999). A Data Partitioning Scheme for Spatial Regression. *International Joint Conference on Neural Networks (IJCNN), vol(4)*, pp. 2474 – 2479.

Wanas, N., Auda, G., Kamel, M.S. & Karray, F. (1998). On the Optimal Number of Hidden Nodes in a Neural Network. *IEEE Canadian Conference on Electrical and Computer Engineering, 1998, vol(2)*, pp. 918 – 921.

Yeung, D. S., Cloete, I., Shi, D. & Ng, W. W. Y. (2010). *Sensitivity Analysis for Neural Network.* London: Springer.

Zhu, Y., Wang, Z., Zhou, J. & Wang, Zh. (2010). Bacteria Classification Using Neural Network. *Sixth International Conference on Natural Computation (ICNC 2010)*, pp. 1199 – 1203.